

Distributions of missing sums and differences

Tiffany C. Inglis

University of British Columbia & University of Waterloo

NSERC USRA Report

August 28, 2007

piffany@gmail.com

1 Introduction

For every set of integers $R = \{x_1, \dots, x_r\}$, there is a corresponding sumset with all pairwise sums

$$R + R := \{x_i + x_j : 1 \leq i, j \leq r\} \quad (1)$$

and a difference set with all possible pairwise differences

$$R - R := \{x_i - x_j : 1 \leq i, j \leq r\}. \quad (2)$$

Various properties of these sets are of interest to us, in particular, when R is taken to be a random subset of $\{0, \dots, n-1\}$ for some positive integer n . Here we introduce some terminologies to be used throughout the paper.

Definition 1.1. For every $n \in \mathbb{N}$, let R_n be the random variable denoting a uniformly randomly chosen subset of $\{0, \dots, n-1\}$ with $P[i \in R_n] = \frac{1}{2}$ for all i satisfying $0 \leq i \leq n-1$. Define R'_n and R''_n similarly with the extra conditions that $0 \in R'_n$ and $\{0, n-1\} \subset R''_n$. A subscript of ∞ indicates that the set is taken from all nonnegative integers.

2 Probability of n as a sum given that $n-1$ is a sum

Associated with each integer $k \geq 0$ is the probability of it being found in $R_\infty + R_\infty$. Martin and O'Bryant [2006] derived an explicit form for this probability:

$$P[k \in R_\infty + R_\infty] = \begin{cases} 1 - \left(\frac{3}{4}\right)^{(k+1)/2}, & \text{if } k \text{ is odd} \\ 1 - \frac{1}{2} \left(\frac{3}{4}\right)^{k/2}, & \text{if } k \text{ is even.} \end{cases} \quad (3)$$

In the following theorem, we extend this notion to find the probability of k being a sum given that $k-1$ is a sum.

Theorem 2.1. For every integer $k \geq 1$, the conditional probability

$$P[k \in R_\infty + R_\infty | k-1 \in R_\infty + R_\infty] = 2 + \frac{F_{k+2} + (-1)^k 3^{\lfloor k/2 \rfloor} - 2^{k+1}}{2(2^k - 3^{\lfloor k/2 \rfloor})}, \quad (4)$$

where F_{k+2} is the $(k+2)$ -th term in the Fibonacci sequence, given explicitly as

$$F_{k+2} = \frac{(1 + \sqrt{5})^{k+2} - (1 - \sqrt{5})^{k+2}}{2^{k+2}\sqrt{5}}. \quad (5)$$

Proof. As the actual derivation is rather tedious, I will spare the details and simply outline the proof. By definition of conditional probability,

$$P[k \in R_\infty + R_\infty | k-1 \in R_\infty + R_\infty] = \frac{P[\{k-1, k\} \subset R_\infty + R_\infty]}{P[k-1 \in R_\infty + R_\infty]}, \quad (6)$$

for which the denominator is easily determined. As for the numerator, let us introduce some new notations. For i satisfying $0 \leq i \leq \frac{k}{2}$, let E_i^k be the event that

$$\{i, k-i\} \subset R_\infty \vee \{i+1, k-(i+1)\} \subset R_\infty \vee \dots \vee \left\{\left\lfloor \frac{k}{2} \right\rfloor, k - \left\lfloor \frac{k}{2} \right\rfloor\right\} \subset R_\infty, \quad (7)$$

and for i satisfying $0 \leq i \leq \frac{k-1}{2}$, let E_i^{k-1} be the event that

$$\{i, (k-1)-i\} \subset R_\infty \vee \{i+1, (k-1)-(i+1)\} \subset R_\infty \vee \dots \vee \left\{\left\lfloor \frac{k-1}{2} \right\rfloor, (n-1) - \left\lfloor \frac{k-1}{2} \right\rfloor\right\} \subset R_\infty. \quad (8)$$

Hence the numerator in Equation 6 may be re-expressed as $P[E_0^{k-1} \wedge E_0^k]$. It turns out that, with proper conditioning of terms, we can write down a recursive formula for probabilities of this form:

$$P[E_i^{k-1} \wedge E_i^k] = \frac{1}{2}P[E_{i+1}^{k-1} \wedge E_{i+1}^k] + \frac{1}{8}P[E_{i+1}^{k-1} \wedge E_{i+2}^k] + \dots \quad (9)$$

$$P[E_i^{k-1} \wedge E_{i+1}^k] = \frac{1}{2}P[E_{i+1}^{k-1} \wedge E_{i+1}^k] + \frac{1}{4}P[E_{i+1}^{k-1} \wedge E_{i+2}^k] + \dots \quad (10)$$

Now if we denote $P[E_i^{k-1} \wedge E_i^k]$ and $P[E_i^{k-1} \wedge E_{i+1}^k]$ respectively by $P_{2i}(k)$ and $P_{2i+1}(k)$, we have the recursive sequence $\{P_i(k)\}_{0 \leq i \leq k-1}$ with $P_{k-1}(k) = P_{k-2}(k) = \frac{1}{4}$ and

$$P_i(k) = \frac{1}{2}P_{i+1}(k) + \frac{1}{4}P_{i+2}(k) - \frac{1}{24} \left(3 + (-1)^{k-i}\right) \left(\frac{3}{4}\right)^{\lfloor (k-i)/2 \rfloor} + \frac{1}{4} \quad \text{for } i = 0, 1, \dots, k-3. \quad (11)$$

Noticing that the recursion depends on $k-i$ rather than on k and i independently prompts us to introduce yet another sequence $T_j := P_{k-j}(k) = P_{k-j-i}(k-1) = \dots = P_0(j)$ with $T_1 = T_2 = \frac{1}{4}$. For $k \geq 3$, the recursive relation works out to be

$$T_k = 1 + \frac{F_{k+2}}{2^{k+1}} - \frac{3^{\lfloor k/2 \rfloor} (4 - (-1)^k)}{2^{k+1}}, \quad (12)$$

where $\{F_i\}$ is the Fibonacci sequence. From this formula yields the desired probability since

$$P[k \in R_\infty + R_\infty | k-1 \in R_\infty + R_\infty] = \frac{T_0}{P[k-1 \in R_\infty + R_\infty]} \quad (13)$$

$$= \frac{T_0}{1 - \frac{1}{4} (3 + (-1)^k) \left(\frac{3}{4}\right)^{\lfloor k/2 \rfloor}}. \quad (14)$$

□

3 Relating several distributions for number of missing sums

Knowing the probability $P[k \in R_\infty + R_\infty | k-1 \in R_\infty + R_\infty]$ is only the tip of the iceberg. What we are really aiming to find out is, for any $n \geq 0$, how $R_n + R_n$ varies in size. Certainly, $\# \{R_n + R_n\}$ varies depending on n , so instead we turn our focus to the number of sums it misses in the range of all possible sums, $\{0, \dots, 2n-2\}$. For convenience, we will incorporate a few new notations. For any integer set R and real interval $[a, b]$, define $f_{[a,b]}^+(R)$ as the number of sums $R + R$ misses in the interval $[a, b]$. Similarly, let $f_{[a,b]}^-(R)$ be the number of differences $R - R$ misses in $[a, b]$. More precisely,

$$f_{[a,b]}^+(R) := \# \{k \in \mathbb{Z} \cap [a, b] : k \notin R + R\}, \quad (15)$$

$$f_{[a,b]}^-(R) := \# \{k \in \mathbb{Z} \cap [a, b] : k \notin R - R\}. \quad (16)$$

For the three *special* sets R_n , R'_n and R''_n defined earlier, we are primarily concerned with the sums they miss in the intervals $[0, n-1]$ and $[0, 2n-2]$, along with the differences missed in $[0, n-1]$ and $[-(n-1), n-1]$. Note that the differences missed in $[-(n-1), n-1]$ are exactly those missed in $[0, n-1]$, plus their negative counterparts.

Martin and O'Bryant [2006] studied several of these probabilities, including $P[f_{[0,2n-2]}^+(R_n) = m]$, the probability of $R_n + R_n$ missing exactly m sums in $[0, 2n-2]$. From randomly generated data, they speculated that underneath this distribution lies a more fundamental distribution involving $P[f_{[0,n-1]}^+(R_n) = m]$, which in turn is built upon the distribution for $P[f_{[0,n-1]}^+(R'_n) = m]$ (Figure 1). The following theorem states the exact relationship between these distributions in the limiting case as n approaches infinity:

Theorem 3.1. *Let m be a nonnegative integer. Then we have*

$$(a) \lim_{n \rightarrow \infty} P[f_{[0,2n-2]}^+(R_n) = m] = \sum_{i=0}^m \lim_{n \rightarrow \infty} P[f_{[0,n-1]}^+(R_n) = i] \lim_{n \rightarrow \infty} P[f_{[0,n-1]}^+(R_n) = m-i],$$

$$(b) \lim_{n \rightarrow \infty} P[f_{[0,n-1]}^+(R_n) = m] = \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} \lim_{n \rightarrow \infty} P[f_{[0,n-1]}^+(R'_n) = m-2i].$$

The proof involves counting the missing sums separately, for example, splitting them into small sums (ranging from 0 to $n-1$) and large sums (ranging from n to $2n-2$). For written simplicity, we employ the following subscripts to denote the intervals involved (the possible parameters R_n and R'_n are omitted here):

$$\begin{aligned} (a) f_{all}^+ &:= f_{[0,2n-2]} & (b) f_S^+ &:= f_{[0,n-1]} & (c) f_L^+ &:= f_{[n,2n-2]} \\ (d) f_{XS}^+ &:= f_{[0, \lfloor n/8 \rfloor - 1]} & (e) f_{MS}^+ &:= f_{[\lfloor n/8 \rfloor, n-1]} & (f) f_{ML}^+ &:= f_{[n, (2n-2) - \lfloor n/8 \rfloor]} \\ (g) f_{XL}^+ &:= f_{[(2n-2) - \lfloor n/8 \rfloor + 1, 2n-2]} & (h) f_M^+ &:= f_{[\lfloor n/8 \rfloor, (2n-2) - \lfloor n/8 \rfloor]} \end{aligned}$$

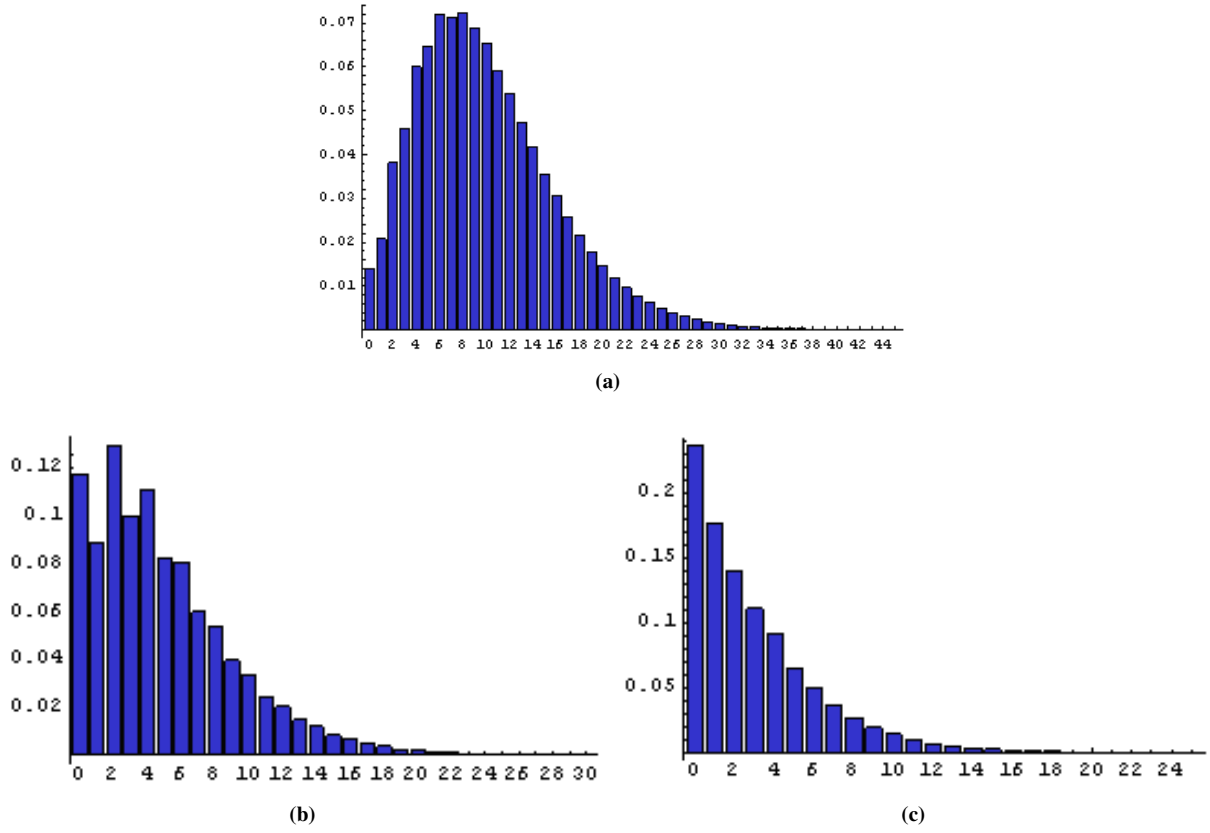


Figure 1: The distributions for the number of missing sums for (a) R_n in $[0, 2n - 2]$, (b) R_n in $[0, n - 1]$, and (c) R'_n in $[0, n - 1]$.

Here are a few lemmas to facilitate the proof of the theorem:

Lemma 3.2.

$$\lim_{n \rightarrow \infty} P[f_M^+(R_n) \geq 1] = 0. \quad (17)$$

Proof. From Equation 3, we can derive the inequality $P[k \notin R_n + R_n] \leq \left(\frac{3}{4}\right)^{\lfloor k/2 \rfloor}$, so

$$\lim_{n \rightarrow \infty} P[f_M^+(R_n) \geq 1] \leq \lim_{n \rightarrow \infty} \sum_{k=\lfloor n/8 \rfloor}^{(2n-2)-\lfloor n/8 \rfloor} P[k \notin R_n + R_n] \quad (18)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{k=\lfloor n/8 \rfloor}^{(2n-2)-\lfloor n/8 \rfloor} \left(\frac{3}{4}\right)^{\lfloor k/2 \rfloor} \quad (19)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{k=\lfloor n/8 \rfloor}^{(2n-2)-\lfloor n/8 \rfloor} \left(\frac{3}{4}\right)^{\lfloor \lfloor n/8 \rfloor / 2 \rfloor} \quad (20)$$

$$\leq \lim_{n \rightarrow \infty} (2n - 2\lfloor n/8 \rfloor - 1) \left(\frac{3}{4}\right)^{\lfloor \lfloor n/8 \rfloor / 2 \rfloor} \quad (21)$$

$$= 0, \quad (22)$$

since the exponential decay of $\left(\frac{3}{4}\right)^{\lfloor \lfloor n/8 \rfloor / 2 \rfloor}$ exceeds the linear growth of $2n - 2\lfloor \frac{n}{8} \rfloor - 1$. A probability is always nonnegative, hence $\lim_{n \rightarrow \infty} P[f_M^+(R_n) \geq 1]$ must equal zero. \square

Lemma 3.3. For any integer $m \geq 0$, the limit of the probability

$$\lim_{n \rightarrow \infty} \sum_{a, c \geq 0, b \geq 1, a+b+c=m} P[f_{XS}^+(R_n) = a \wedge f_M^+(R_n) = b \wedge f_{XL}^+(R_n) = c] = 0. \quad (23)$$

Proof. Using Lemma 3.2, we have

$$\lim_{n \rightarrow \infty} \sum_{a, c \geq 0, b \geq 1, a+b+c=m} P[f_{XS}^+(R_n) = a \wedge f_M^+(R_n) = b \wedge f_{XL}^+(R_n) = c] \leq \lim_{n \rightarrow \infty} \sum_{b \geq 1} P[f_M^+(R_n) = b] \quad (24)$$

$$\leq \lim_{n \rightarrow \infty} P[f_M^+(R_n) \geq 1] \quad (25)$$

$$= 0. \quad (26)$$

□

Lemma 3.4. For any integer m satisfying $m \geq 0$, the probability

$$P[f_{XS}^+(R_n) = i \wedge f_{XL}^+(R_n) = m - i] = P[f_{XS}^+(R_n) = i] P[f_{XL}^+(R_n) = m - i]. \quad (27)$$

Proof. To prove the lemma, we need to show that $f_{XS}^+(R_n) = i$ and $f_{XL}^+(R_n) = m - i$ are independent events. By definition,

$$f_{XS}^+(R_n) := f_{[0, \lfloor n/8 \rfloor - 1]}(R_n) \quad (28)$$

$$= \frac{\lfloor \frac{n}{8} \rfloor - \# \{ \{0, \dots, \lfloor \frac{n}{8} \rfloor - 1\} \cap (R_n + R_n) \}}{\lfloor \frac{n}{8} \rfloor} \quad (29)$$

$$= \frac{\lfloor \frac{n}{8} \rfloor - \# \{ (\{0, \dots, \lfloor \frac{n}{8} \rfloor - 1\} \cap R_n) + (\{0, \dots, \lfloor \frac{n}{8} \rfloor - 1\} \cap R_n) \}}{\lfloor \frac{n}{8} \rfloor}, \quad (30)$$

where the last step uses the fact that any sum of R_n in the range $[0, \lfloor \frac{n}{8} \rfloor - 1]$ can only be the sum of two elements in R_n in that range. Similarly,

$$f_{XL}^+(R_n) = \frac{\lfloor \frac{n}{8} \rfloor - \# \{ (\{2n - \lfloor \frac{n}{8} \rfloor - 1, \dots, 2n - 2\} \cap R_n) + (\{2n - \lfloor \frac{n}{8} \rfloor - 1, \dots, 2n - 2\} \cap R_n) \}}{\lfloor \frac{n}{8} \rfloor}, \quad (31)$$

since large sums only result from adding two large elements together.

Notice that $f_{XS}^+(R_n)$ is a function of n and $\{0, \dots, \lfloor \frac{n}{8} \rfloor - 1\} \cap R_n$ while $f_{XL}^+(R_n)$ is a function of n and $\{2n - \lfloor \frac{n}{8} \rfloor - 1, \dots, 2n - 2\} \cap R_n$. Since $\{0, \dots, \lfloor \frac{n}{8} \rfloor - 1\} \cap R_n$ and $\{2n - \lfloor \frac{n}{8} \rfloor - 1, \dots, 2n - 2\} \cap R_n$ are disjoint, for any fixed n , the values of $f_{XS}^+(R_n)$ and $f_{XL}^+(R_n)$ are independent. □

Lemma 3.5. For any integer m satisfying $m \geq 0$, the limit of the probability

$$\lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+(R_n) = i] P[f_{XL}^+(R_n) = m - i] = \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_S^+(R_n) = i] P[f_L^+(R_n) = m - i]. \quad (32)$$

Proof. To prove the statement, we will show both

$$\lim_{n \rightarrow \infty} P[f_{XS}^+(R_n) = i] = \lim_{n \rightarrow \infty} P[f_S^+(R_n) = i] \quad (33)$$

$$\lim_{n \rightarrow \infty} P[f_{XL}^+(R_n) = m - i] = \lim_{n \rightarrow \infty} P[f_L^+(R_n) = m - i]. \quad (34)$$

Re-express the probability

$$P[f_S^+(R_n) = i] = P[f_{XS}^+(R_n) = i \wedge f_{MS}^+(R_n) = 0] + P[f_{XS}^+(R_n) = i - f_M^+(R_n) \wedge f_{MS}^+(R_n) \geq 1] \quad (35)$$

$$= P[f_{XS}^+(R_n) = i \wedge f_{MS}^+(R_n) = 0] + \mathcal{O}(P[f_{MS}^+(R_n) \geq 1]) \quad (36)$$

$$= P[f_{XS}^+(R_n) = i] - P[f_{XS}^+(R_n) = i \wedge f_{MS}^+(R_n) \geq 1] + \mathcal{O}(P[f_{MS}^+(R_n) \geq 1]) \quad (37)$$

$$= P[f_{XS}^+(R_n) = i] + \mathcal{O}(P[f_{MS}^+(R_n) \geq 1]) + \mathcal{O}(P[f_{MS}^+(R_n) \geq 1]) \quad (38)$$

$$= P[f_{XS}^+(R_n) = i] + \mathcal{O}(P[f_{MS}^+(R_n) \geq 1]) \quad (39)$$

$$= P[f_{XS}^+(R_n) = i] + \mathcal{O}(P[f_M^+(R_n) \geq 1]). \quad (40)$$

Since $\lim_{n \rightarrow \infty} P[f_M^+(R_n)] \geq 1 = 0$ by Lemma 3.2, we have

$$\lim_{n \rightarrow \infty} P[f_S^+(R_n) = i] = \lim_{n \rightarrow \infty} P[f_{XS}^+(R_n) = i]. \quad (41)$$

Analogously, we can show that

$$\lim_{n \rightarrow \infty} P[f_L^+(R_n) = i] = \lim_{n \rightarrow \infty} P[f_{XL}^+(R_n) = i]. \quad (42)$$

□

Now we are in position to prove Theorem 3.1, starting with Part (a). To avoid redundancy, we will write f^+ in place of $f^+(R_n)$ throughout this proof, combined with various pre-defined subscripts. Now to analyze $P[f_{all}^+ = m]$, first partition f_{all}^+ into f_{XS}^+ , f_M^+ , and f_{XL}^+ , followed by conditioning on whether $f_M^+ = 0$, as shown:

$$\begin{aligned} P[f_{all}^+ = m] &= \sum_{a,b,c \geq 0, a+b+c=m} P[f_{XS}^+ = a \wedge f_M^+ = b \wedge f_{XL}^+ = c] \\ &= \sum_{a,c \geq 0, a+c=m} P[f_{XS}^+ = a \wedge f_M^+ = 0 \wedge f_{XL}^+ = c] + \sum_{a,c \geq 0, b \geq 1, a+b+c=m} P[f_{XS}^+ = a \wedge f_M^+ = b \wedge f_{XL}^+ = c] \\ &= \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ = 0 \wedge f_{XL}^+ = m-i] + \sum_{a,c \geq 0, b \geq 1, a+b+c=m} P[f_{XS}^+ = a \wedge f_M^+ = b \wedge f_{XL}^+ = c]. \end{aligned} \quad (43)$$

Taking the limit as n approaches infinity then applying Lemma 3.2, we have

$$\lim_{n \rightarrow \infty} P[f_{all}^+ = m] = \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ = 0 \wedge f_{XL}^+ = m-i]. \quad (46)$$

The RHS of the equation can be rewritten as

$$\lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_{XL}^+ = m-i] - \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ \geq 1 \wedge f_{XL}^+ = m-i]. \quad (47)$$

Taking the absolute value of Equation 47, we get

$$\left| \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_{XL}^+ = m-i] - \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ \geq 1 \wedge f_{XL}^+ = m-i] \right| \quad (48)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ \geq 1 \wedge f_{XL}^+ = m-i] \quad (49)$$

$$\leq \lim_{n \rightarrow \infty} P[f_M^+ \geq 1] = 0, \text{ by Lemma 3.2.} \quad (50)$$

So the expression in Equation 47 equals zero. In other words,

$$\lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_M^+ = 0 \wedge f_{XL}^+ = m-i] = \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_{XL}^+ = m-i], \quad (51)$$

which implies $\lim_{n \rightarrow \infty} P[f_{all}^+ = m] = \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i \wedge f_{XL}^+ = m-i]$. Then by Lemma 3.4 and 3.5,

$$\lim_{n \rightarrow \infty} P[f_{all}^+ = m] = \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_{XS}^+ = i] P[f_{XL}^+ = m-i] \quad (52)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_S^+ = i] P[f_L^+ = m-i], \text{ by symmetry of } f_S \text{ and } f_L, \quad (53)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=0}^m P[f_S^+ = i] P[f_S^+ = m-i], \text{ as desired.} \quad (54)$$

Now for Part (b), begin by conditioning $P[f_S^+(R_n) = m]$ on whether the sumset $R_n + R_n$ is empty, which yields

$$P[R_n + R_n = \emptyset] P[f_S^+(R_n) = m | R_n + R_n = \emptyset] + P[R_n + R_n \neq \emptyset] P[f_S^+(R_n) = m | R_n + R_n \neq \emptyset]. \quad (55)$$

The first term in Equation 55 equals $P[R_n \neq \emptyset] P[f_S^+(R_n) = m | f_S^+(R_n) = n] = \left(\frac{1}{2^n}\right) \cdot 1_{[m=n]}$. Hence

$$\lim_{n \rightarrow \infty} P[R_n + R_n = \emptyset] P[f_S^+(R_n) = m | R_n + R_n = \emptyset] = 0. \quad (56)$$

Then condition the second term in Equation 55 on the value of the minimum sum, $\min(R_n + R_n)$:

$$P[R_n + R_n \neq \emptyset] P[f_{[0, n-1]}^+(R_n) = m | R_n + R_n \neq \emptyset] \quad (57)$$

$$= \sum_{i=0}^{2n-2} P[\min(R_n + R_n) = i] P[f_{[0, n-1]}^+(R_n) = m | \min(R_n + R_n) = i] \quad (58)$$

$$= \sum_{i=0}^m P[\min(R_n + R_n) = i] P[f_{[0, n-1]}^+(R_n) = m | \min(R_n + R_n) = i] \quad (59)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} P[\min(R_n) = i] P[f_{[0, n-1]}^+(R_n) = m | \min(R_n) = i] \quad (60)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-1]}^+(R_n) = m | \min(R_n) = i] \quad (61)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-1]}^+(R'_{n-i} + i) = m] \quad (62)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[2i, n-1]}^+(R'_{n-i} + i) = m - 2i] \quad (63)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-2i-1]}^+(R'_{n-i}) = m - 2i]. \quad (64)$$

$$(65)$$

Express this result as

$$\sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) - f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) = m - 2i] \quad (66)$$

$$= \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) = 0] - \quad (67)$$

$$- \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1]. \quad (68)$$

In Equation 68, the limit of the second term

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1] \quad (69)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1] \quad (70)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[n-2i-1, n-i-1]}^+(R_{n-i}) \geq 1], \text{ and for } n \gg m, \quad (71)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_M^+(R_{n-i}) \geq 1] = 0, \text{ by Lemma 3.2.} \quad (72)$$

Hence, $\lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1] = 0$. We have just shown that $P[f_S^+(R_n) = m]$

expands into three terms, two of which converge to 0 as $n \rightarrow \infty$, leaving us

$$\lim_{n \rightarrow \infty} P[f_S^+(R_n) = m] = \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[0, n-i-1]}^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) = 0] \quad (73)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_S^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) = 0]. \quad (74)$$

By the above equation, the difference

$$\left| \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_S^+(R'_{n-i}) = m - 2i] - \lim_{n \rightarrow \infty} P[f_S^+(R_n) = m] \right| \quad (75)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_S^+(R'_{n-i}) = m - 2i \wedge f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1] \quad (76)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[n-2i-1, n-i-1]}^+(R'_{n-i}) \geq 1] \quad (77)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_{[n-2i-1, n-i-1]}^+(R_{n-i}) \geq 1], \text{ and for } n \gg m, \quad (78)$$

$$\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_M^+(R_{n-i}) \geq 1], \text{ by Lemma 3.2.} \quad (79)$$

$$(80)$$

Hence,

$$\lim_{n \rightarrow \infty} P[f_S^+(R_n) = m] = \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_S^+(R'_{n-i}) = m - 2i] \quad (81)$$

$$= \lim_{n \rightarrow \infty} \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} P[f_S^+(R'_n) = m - 2i], \text{ as required.} \quad (82)$$

In fact, with this method, we can extend Theorem 3.1 as follows (proof omitted):

Theorem 3.6. *Let m be a nonnegative integer. Then we have*

$$(a) \lim_{n \rightarrow \infty} P[f_1 = m] = \sum_{i=0}^m \lim_{n \rightarrow \infty} P[f_2 = i] \lim_{n \rightarrow \infty} P[f_2 = m - i],$$

$$\text{where } (f_1, f_2) \in \left\{ \left(f_{[0, 2n-2]}^+(R_n), f_{[0, n-1]}^+(R_n) \right), \left(f_{[0, 2n-2]}^+(R'_n), f_{[0, n-1]}^+(R'_n) \right) \right\}$$

$$(b) \lim_{n \rightarrow \infty} P[f_3 = m] = \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{1}{2^{i+1}} \lim_{n \rightarrow \infty} P[f_4 = m - 2i],$$

$$\text{where } (f_3, f_4) \in \left\{ \left(f_{[0, n-1]}^+(R_n), f_{[0, n-1]}^+(R'_n) \right), \left(f_{[0, 2n-2]}^+(R_n), f_{[0, 2n-2]}^+(R'_n) \right), \left(f_{[0, 2n-2]}^+(R'_n), f_{[0, 2n-2]}^+(R''_n) \right), \right. \\ \left. \left(f_{[0, 2n-2]}^-(R_n), f_{[0, 2n-2]}^-(R'_n) \right), \left(f_{[0, 2n-2]}^-(R'_n), f_{[0, 2n-2]}^-(R''_n) \right) \right\}.$$

4 Missing fewer sums and differences is more probable?

Now that we know how the distributions of number of missing sums and differences are related, let us turn our attention to the properties of each one. The most fundamental distributions for sums and differences are $P[f_{[0, n-1]}^+(R'_n) = m]$ and $P[f_{[0, n-1]}^-(R''_n) = m]$, shown in Figure 2.

While both appear to be decreasing at a roughly exponential rate, the tiny distribution actually has a tiny *blip* in the tail. We will state this as a theorem:

Theorem 4.1. *For $m \geq 0$, $P[f_{[0, n-1]}^-(R''_n) = m]$ is not decreasing in m except for $n = 1, 2, 3, 5, 9$.*

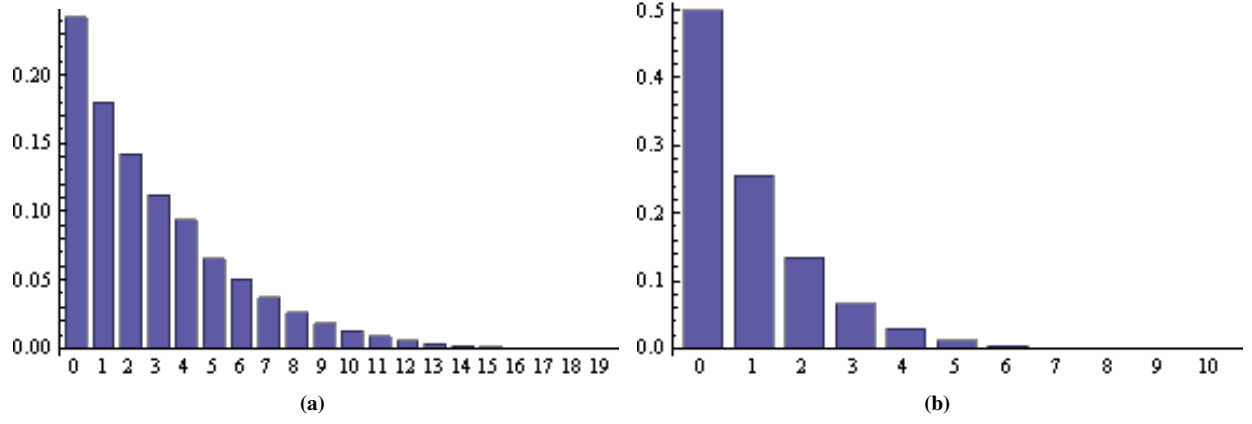


Figure 2: The observed distributions of (a) the number of missing sums for R'_n in $[0, n-1]$ and (b) the number of missing differences for R''_n in $[0, n-1]$.

Proof. Let $p_n(m) = P[f_{[0, n-1]}^-(R''_n) = m]$. For $n \leq 10$, we can easily verify the claim by determining the exact function. Figure 3 shows whether or not these functions are decreasing.

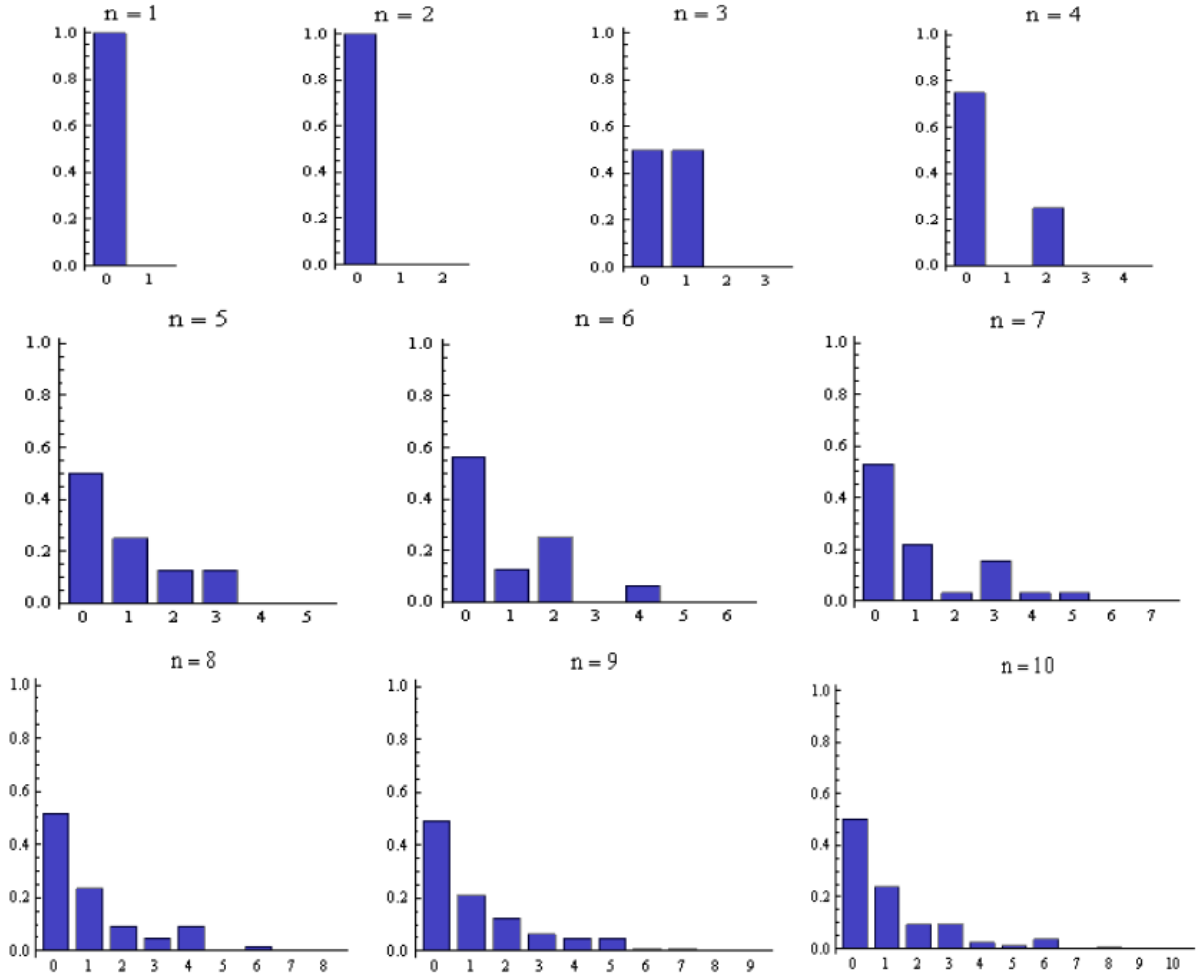


Figure 3: Exact distributions of the number of missing differences for small values of n

To prove the theorem for $n \geq 11$, we want to find some $m \in \{1, \dots, n\}$ such that

$$p_n(m) - p_n(m-1) > 0. \quad (83)$$

Let us evaluate $p_n(m)$ starting with $m = n, n-1, n-2, n-3, \dots$

- $p_n(n) = P[R_n'' - R_n'' = \emptyset] = 0$, since $\{0, n-1\} \subset R_n''$ implies $\{0, n-1\} \subset R_n'' - R_n''$
- $p_n(n-1) = 0$
- $p_n(n-2) = P[R_n'' = \{0, n-1\}] = \frac{1}{2^{n-2}}$
- $p_n(n-3) = P[R_n'' = \{0, \frac{n-1}{2}, n-1\}] = \left(\frac{1}{2^{n-2}}\right) \cdot 1_{[2|n-1]}$
- $p_n(n-4) = P[R_n'' = \{0, a, n-1\} | a \neq \frac{n-1}{2}] + P[R_n'' = \{0, \frac{1}{3}(n-1), \frac{2}{3}(n-1), n-1\}]$
 $= \frac{n-2}{2^{n-2}} - \left(\frac{1}{2^{n-2}}\right) \cdot 1_{[2|n-1]} + \left(\frac{1}{2^{n-2}}\right) \cdot 1_{[3|n-1]}$
- $p_n(n-5) = P[R_n'' = \{0, d, (n-1) - d, n-1\} | d \neq \frac{1}{3}(n-1)] +$
 $+ P[R_n'' = \{0, \frac{1}{4}(n-1), \frac{1}{2}(n-1), (n-1)\} | d_1 = d_2 = \frac{1}{2}d_3] +$
 $+ P[R_n'' = \{0, \frac{1}{2}(n-1), \frac{3}{4}(n-1), (n-1)\} | \frac{1}{2}d_1 = d_2 = d_3] +$
 $+ P[R_n'' = \{0, \frac{1}{4}(n-1), \frac{1}{2}(n-1), \frac{3}{4}(n-1), n-1\}] +$
 $= \frac{\left[\frac{n-1}{2}\right] - 1}{2^{n-2}} - \left(\frac{1}{2^{n-2}}\right) \cdot 1_{[3|n-1]} + \left(\frac{3}{2^{n-2}}\right) \cdot 1_{[4|n-1]}.$

So in the case of $m = n-4$,

$$p_n(m) - p_n(m-1) = p_n(n-4) - p_n(n-5) \quad (84)$$

$$= \frac{n-2}{2^{n-2}} - \left(\frac{1}{2^{n-2}}\right) 1_{[2|n-1]} + \left(\frac{1}{2^{n-2}}\right) 1_{[3|n-1]} - \frac{\left[\frac{n-1}{2}\right] - 1}{2^{n-2}} + \left(\frac{1}{2^{n-2}}\right) 1_{[3|n-1]} - \left(\frac{3}{2^{n-2}}\right) 1_{[4|n-1]} \quad (85)$$

$$\geq \frac{n-2}{2^{n-2}} - \frac{1}{2^{n-2}} - \frac{\left[\frac{n-1}{2}\right] - 1}{2^{n-2}} - \frac{3}{2^{n-2}} \quad (86)$$

$$\geq \frac{n-5 - \frac{n}{2}}{2^{n-2}} \quad (87)$$

$$= \frac{n-10}{2^{n-1}} > 0. \quad (88)$$

□

Using the same idea of analyzing the tail-end distribution, we can extend Theorem 4.2 to include several distributions that seem to be decreasing after a certain point (usually the maximum) but in fact are not due to small blips in the tails.

Theorem 4.2. *In the following distributions, the function reaches a maximum before exhibiting a decreasing trend. For each distribution, let m^* be where the maximum is reached, then we have*

(a) For $m \geq m^*$, $P[f_{[0, n-1]}^-(R_n'') = m]$ is not decreasing except for $n = 1, 2, 3, 5, 9$.

(b) For $m \geq m^*$, $P[f_{[0, n-1]}^-(R_n') = m]$ is not decreasing except for $n = 1, 2$.

(c) For $m \geq m^*$, $P[f_{[0, n-1]}^-(R_n) = m]$ is not decreasing except for $n = 1$.

(d) For $m \geq m^*$, $P[f_{[0, 2n-2]}^+(R_n'') = m]$ is not decreasing except for $n = 1, 2$.

(e) For $m \geq m^*$, $P[f_{[0, 2n-2]}^+(R_n') = m]$ is not decreasing except for $n = 1$.

(f) For $m \geq m^*$, $P[f_{[0, 2n-2]}^+(R_n) = m]$ is not decreasing except for $n = 1$.

For large enough values of n , these distributions stabilize and reach a limit, as shown in Figure 4.

Basically, the theorem says that, even in the parts of the graphs that exhibit a downward trend, these functions are not strictly decreasing due to some blips in the tail-end. Notice that the theorem does not assert anything for the limiting distributions because it is very possible for the limiting function to be decreasing, as in the case of $P[f_{[0, n-1]}^-(R_n'') = m]$, for example. Another interesting observation is that the functions dealing with the number of missing sums in the interval $[0, n-1]$ may be decreasing. We will formally state this as a conjecture:

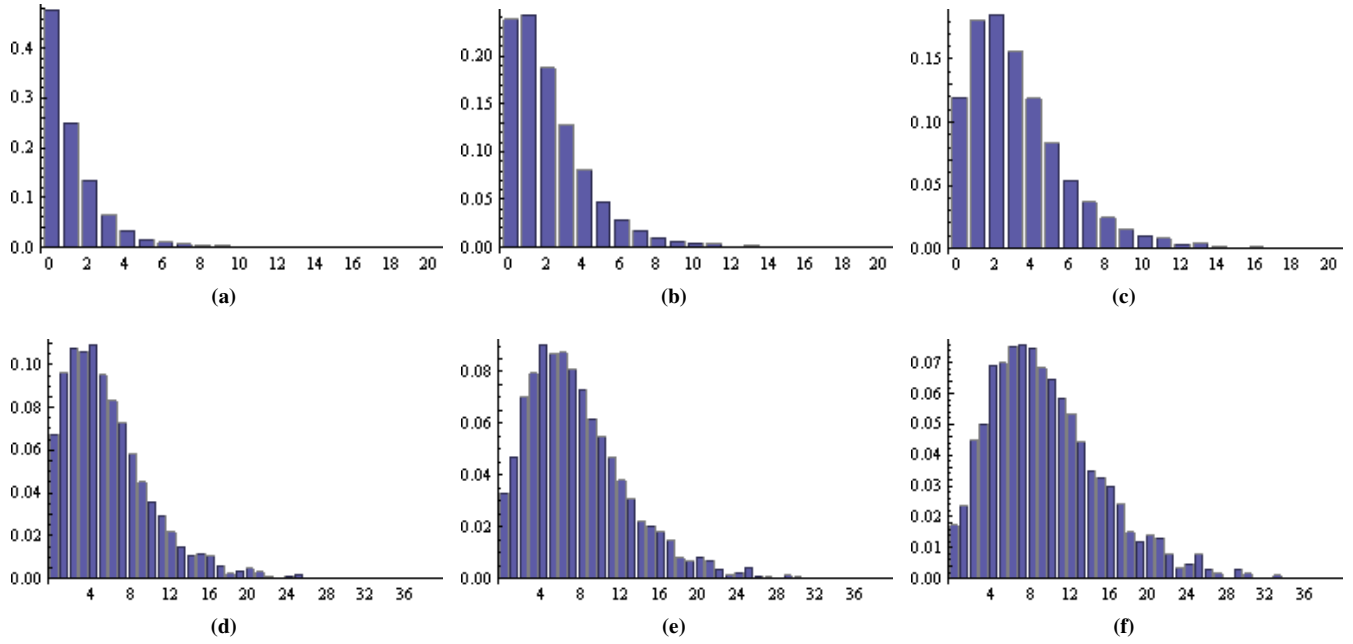


Figure 4: Non-decreasing distributions corresponding to Theorem 4.2

Conjecture 4.1. For each distribution, let m^* be where the maximum is reached, then we have

- (a) For $m \geq m^*$, $P[f_{[0,n-1]}^+(R''_n) = m]$ is decreasing.
- (b) For $m \geq m^*$, $P[f_{[0,n-1]}^+(R'_n) = m]$ is decreasing except for $n = 4, 5, 6$.
- (c) For $m \geq m^*$, $P[f_{[0,n-1]}^+(R_n) = m]$ is decreasing except for $n = 1$.

These claims are supported by Figure 5.

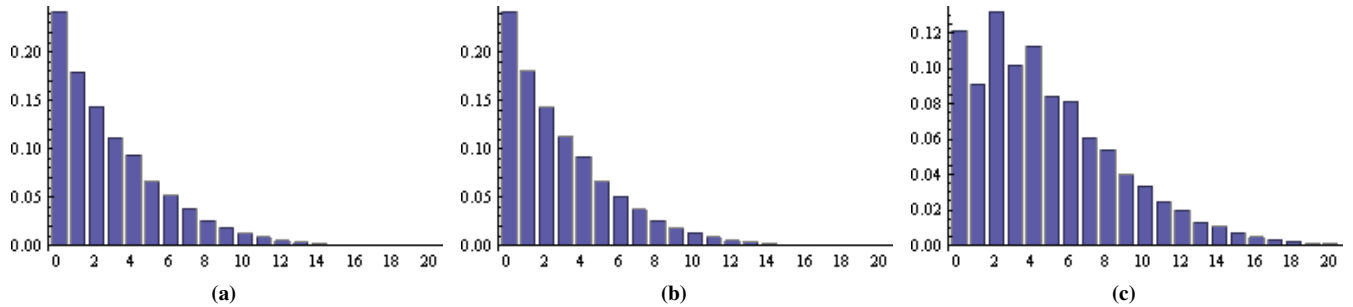


Figure 5: Distributions that appear to be decreasing as described in Conjecture 4.1

5 A conjecture about the limiting distribution for differences

Recall that $P[f_{[0,n-1]}^-(R''_\infty) = m]$ from Figure 4a and $P[f_{[0,n-1]}^+(R'_\infty) = m]$ from Figure 5b are the two fundamental distributions, but despite our efforts, we still do not know much about either except some decreasing properties from the last section. Presently, we want to find out more about the height of the bars in the graphs through experimentation in Mathematica.

Let us start with counting the sums. Instead of looking at all of R'_∞ , which almost surely spans $[0, \infty)$, we will focus on only the elements of R'_∞ the interval $[0, 19]$. By definition, $0 \in R'_\infty$. As for the other elements in $[1, 19]$, we will experiment with all combinations. Since for each experiment, we fix which elements from $[0, 19]$ are in R'_∞ , we also know exactly which elements from $[0, 19]$ are in $R'_\infty + R'_\infty$ because

elements greater than 19 do not affect sums in between 0 and 19. Counting the missing sums for these sets in the interval $[0, 19]$, we have a distribution as shown in Figure 6a.

Notice the striking similarity between this graph and Figure 5a. This resemblance signifies that very few sets miss more than 20 sums, and the sums missed are almost less than 20. In fact, we can observe this pattern by focusing on an event smaller interval, say $[0, 13]$. Very few missed sums exceed 13.

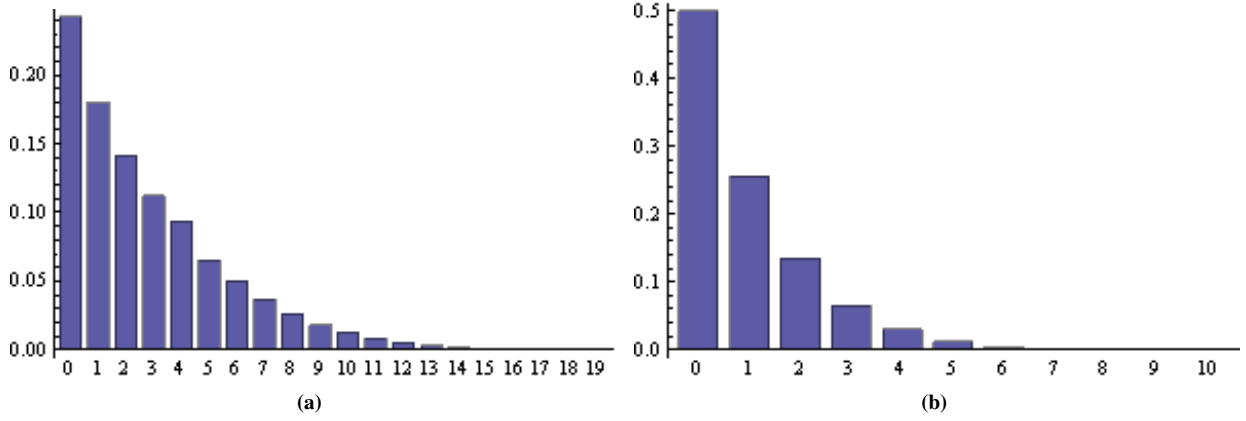


Figure 6: Distributions for (a) the number of missing sums in $[0, 19]$, and (b) the number of missing differences in $[0, 10]$

Likewise, we can take a finite set, insist that both endpoints be included, and iterate through all possible combinations. Counting all relevant pairwise differences, we arrive at a distribution as shown in Figure 6b, which bears an uncanny resemblance to Figure 4a, hence demonstrating that nearly all missed differences are among the 10 largest possible differences. So if d_{max} is the largest possible difference, then $\{d_{max} - 9, \dots, d_{max}\}$ would comprise virtually all the missed differences.

Let us draw our attention to the graph concerning missing differences. Recall that Figure 6b is obtained by iterating through all subsets of $[0, 19]$ containing $\{0, 19\}$ and counting the number of missing differences from $[0, 10]$. Now, for $n = 1, \dots, 9$, we can do the same by taking all subsets of $[0, 2n - 1]$ containing $\{0, 2n - 1\}$ and counting missing differences from $[0, n]$, we get the following distributions as shown in Figure 7.

These graphs suggest that as n approaches ∞ , a limiting distribution exists and, in addition, is the same as the one shown in Figure 4a. Let us state this as a lemma and prove it before proceeding.

Lemma 5.1.

$$\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_n'') = m \right] = \lim_{n \rightarrow \infty} P \left[f_{[n, 2n-1]}^-(R_{2n}'') = m \right]. \quad (89)$$

Proof. We have two known identities. The first

$$\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_n'') = m \right] = \lim_{n \rightarrow \infty} P \left[f_{[0, 2n-1]}^-(R_{2n}'') = m \right]. \quad (90)$$

is obtained by substituting n with $2n - 1$, and the second is

$$\lim_{n \rightarrow \infty} P \left[f_{[0, 2n-1]}^-(R_{2n}'') = m \right] = \sum_{i=0}^m \left(\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_{2n}'') = i \right] \lim_{n \rightarrow \infty} P \left[f_{[n, 2n-1]}^-(R_{2n}'') = m - i \right] \right). \quad (91)$$

So to prove the lemma, we need to show that $\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_{2n}'') = 0 \right] = 1$. Taking the probability of its complement, we have

$$\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_{2n}'') \geq 1 \right] \leq \lim_{n \rightarrow \infty} \sum_{k=0}^n P \left[k \notin R_{2n}'' - R_{2n}'' \right] \quad (92)$$

$$= \lim_{n \rightarrow \infty} \sum_{k=0}^n \left(\frac{3}{4} \right)^{2n-k} \quad (93)$$

$$\leq \lim_{n \rightarrow \infty} (n+1) \left(\frac{3}{4} \right)^n = 0. \quad (94)$$

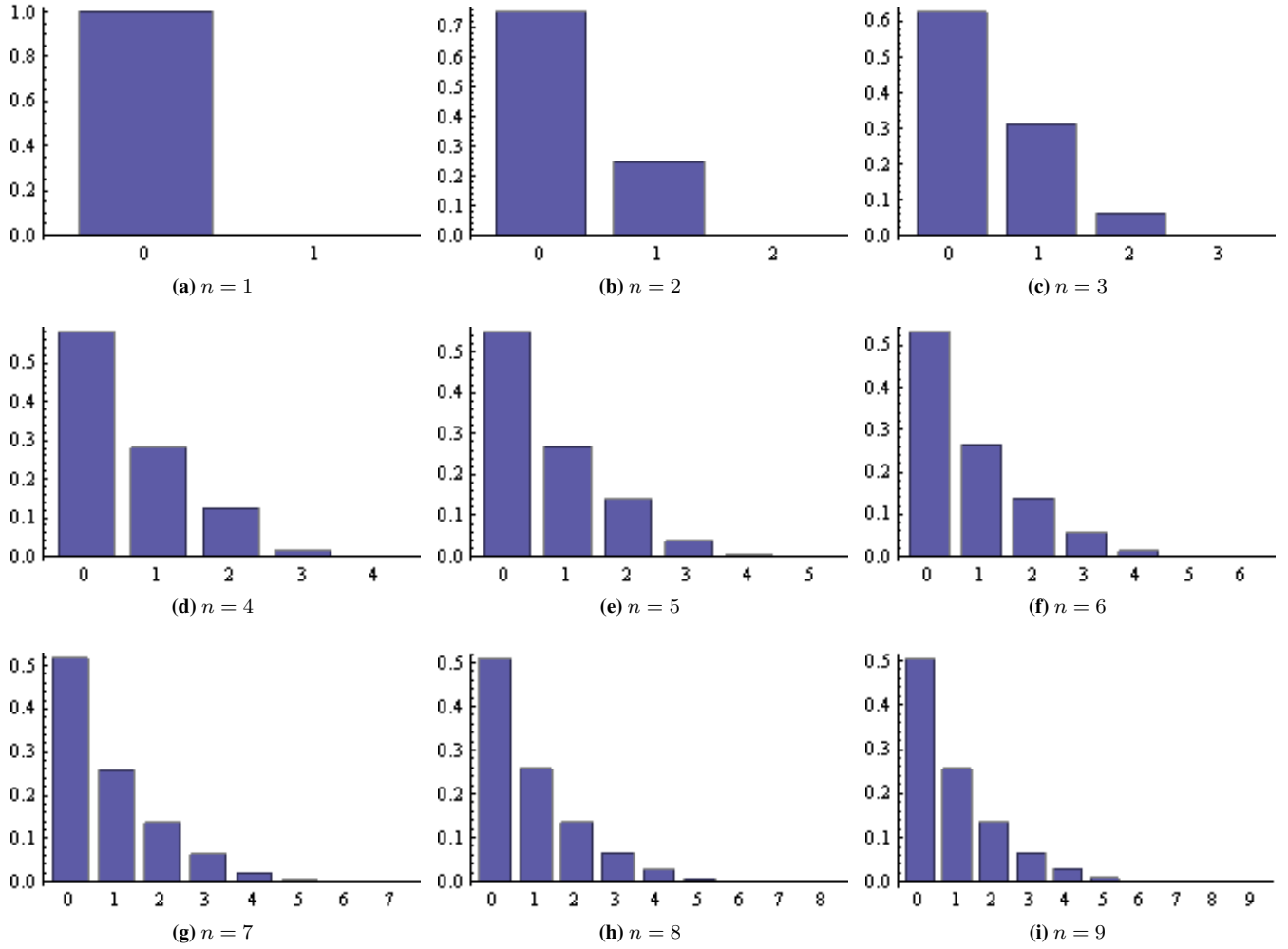


Figure 7: Distribution of the number of missing differences for each value of n

Then since $\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_{2n}'') \geq 1 \right] = 0$, we conclude that

$$\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_n'') = m \right] = \lim_{n \rightarrow \infty} P \left[f_{[0, 2n-1]}^-(R_{2n}'') = m \right] \quad (95)$$

$$= \lim_{n \rightarrow \infty} P \left[f_{[n, 2n-1]}^-(R_{2n}'') = m \right]. \quad (96)$$

□

The significance of all this is that, while $P \left[f_{[0, n-1]}^-(R_n'') = m \right]$ is, in general, not a decreasing function, $P \left[f_{[0, n-1]}^-(R_{2n}'') = m \right]$ *does* seem to be decreasing, according to Theorem 4.2(a). If the decreasing trend continues, then its limiting distribution would also be decreasing. Note that this does not contradict with Theorem 4.2(a) since a sequence of functions does not need to be decreasing for its limit to be decreasing. In the proof of the theorem, we showed that a small blip at the tail-end of the distribution prevents it from being decreasing. If the blip is the only anomaly in an otherwise decreasing function, then as n approaches ∞ , the height of the blip would decrease to 0, resulting in a decreasing limiting distribution. Figure 7 serves to strengthen this last conjecture:

Conjecture 5.1. For all integers $m \geq 0$, $\lim_{n \rightarrow \infty} P \left[f_{[0, n-1]}^-(R_n'') = m \right]$ is a decreasing function.

References

MARTIN, G., AND O'BRYANT, K. 2006. Many sets have more sums than differences. arXiv:math.NT/0608131v3.